Flexible Docking Simulations: Scaled Collective Variable Monte Carlo Minimization Approach Using Bezier Splines, and Comparison with a Standard Monte Carlo Algorithm

JEAN-YVES TROSSET, HAROLD A. SCHERAGA

Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301

Received 3 July 1998; accepted 3 October 1998

ABSTRACT: An algorithm for docking a flexible ligand onto a flexible or rigid receptor, using the scaled-collective-variables Monte Carlo with energy minimization approach, is presented. Energy minimization is shown to be one of the best techniques for distinguishing between native- and nonnative-generated conformations. Incorporation of this technique into a Monte Carlo procedure enables one to distinguish the native conformation directly during the conformational search. It avoids the generation of a large number of ligand conformers for which more sophisticated energy evaluation tools would have had to be applied to identify the nativelike conformations. The efficiency of the Monte Carlo minimization was greatly improved by incorporating a new grid-based energy evaluation technique using Bezier splines for which the energy function, as well as all of its derivatives, can be deduced from the values at the grid points. Comparison between our ECEPP/3-based algorithm and the Monte Carlo algorithm presented elsewhere (Hart, T. N.; Read, R. J. Prot Struct Funct Genet 1992, 13, 206–222) has been made for docking NH_2—D—Phe—Pro—Arg—COOH, the noncovalent analog of NH_2—D—Phe—Pro—Arg—chloromethylketone (PPACK), onto the active site of human α-thrombin.
 J Comput Chem 20: 244–252, 1999

Keywords: flexible ligand; flexible or rigid receptor; grid-based energy calculation; ECEPP/3-based algorithm; active site of thrombin; tripeptide ligand

Correspondence to: Harold A. Scheraga
Contract/grant sponsor: NSF; contract grant number: MCB95-13167
Contract/grant sponsor: NIH; contract grant numbers: GM-14312 and HL-30616
Contract/grant sponsor: Cornell Biotechnology Center

 CCC 0192-8651 / 99 / 020244-09

Introduction

Determination of the binding free energy in a ligand–receptor complex is one of the challenging problems in computer simulation. Because of the difficulty of sampling the whole conformational space of the system, the free energy is estimated primarily from a set of relevant conformations; that is, the structures that are close to the native conformation of the complex. If the structure of the receptor is known beforehand, the problem consists of finding the optimum conformation of the ligand that best fits inside the cavity of the active site of the receptor. The search for the optimum conformation is usually based on simple geometric criteria, such as shape complementarity and/or important energy contributions such as hydrogen bonds, hydrophobic interactions, electrostatic interactions, etc., between the ligand and the receptor.[1–13] Whichever geometric or energetic criterion is used for a docking algorithm, one faces the multiple-minima problem in finding the global minimum (GM) of a multivariable cost or energy function. The success of the approach thus depends on two necessary conditions: (i) the optimal value of the cost function should correspond to the native conformation of the complex, as observed from X-ray or NMR experiments; and (ii) the probability of reaching the GM during a multiple-start procedure should be significantly higher (more than a few percent).

Two general approaches are available to solve the docking problem. The first is to use simple geometric criteria and/or simple energy functions to produce a list of conformations that are potential candidates for native conformation. This list is then filtered by using a more precise energy evaluation procedure, such as an electrostatic solvation calculation, an empirical free energy calculation, etc.,[14,15] with the hope that the nativelike conformation will be ranked at the top of the list. The rationale of this approach is to save some computational time by carrying out more sophisticated energy evaluation procedures only on a preselected set of low-energy conformations. The drawback is that a nativelike conformation of the ligand can be rejected during the sampling procedure because of a few atomic clashes between the ligand and the receptor that could easily be relieved by local energy minimization. For complicated binding modes, in which the ligand is highly flexible and has to fit into a narrow pocket, the probability of finding this conformation by a purely random search is very small. One way to increase the probability of success in the random search is to carry out a local energy minimization after every generation of a new random structure. This would quickly remove atomic clashes between the ligand and the receptor. This second approach is the one that has been adopted in this study. Because docking algorithms should meet the requirement of being able to dock a large number of molecules in a relatively short time, this approach has not been used widely.[16–18] However, Shoichet and Kuntz[2] have shown that energy minimization with an all-atom force field is one of the most powerful methods for distinguishing native from nonnative conformations of a ligand–receptor complex. Furthermore, the incorporation of local energy minimization into a conformational search, such as a Monte Carlo[19–21] or a genetic algorithm,[22] has also been shown to be very efficient in finding the GM of polypeptides or protein–ligand complexes.

Recently, the development of a new grid-based energy technique using Bezier splines[23] has made the energy minimization of a ligand–receptor complex very efficient.[21] This arises because most, if not all, of the atoms of the receptor can be considered as rigid, and their contribution to the energetic field in the region of the active site can be stored into a three-dimensional (3D) grid. This changes the scaling of the number of interactions from $N_{ligand} \times N_{receptor}$ to N_{ligand}, thereby increasing the speed of the energy calculation. Until now, the grid technique was used only for energy calculations. The energy value at a given point of the active site was usually estimated by trilinear or cubic interpolation[24] from the surrounding grid points. Gradient-based energy minimizations were not carried out on the grid, because continuous and differentiable approximate energy functions were not obtainable from the grid points. Bezier splines provide a remedy for this technical problem. Full advantage of the grid technique can now be taken for local energy minimization, in addition to the evaluation of the energy itself.

In the present article, we examine the Monte Carlo energy minimization method for docking a flexible ligand onto a rigid receptor. We test our protocol with a tripeptide (FPR), NH_2—D—Phe—Pro—Arg—COOH, which is the noncovalent analog of NH_2—D—Phe—Pro—Arg—chloromethylketone (PPACK), an antagonist of human α-thrombin.[25,26] The questions that we

address in this study are: (1) What is the probability of finding the GM for a highly flexible ligand such as the one used here (with 17 internal degrees of freedom including the dihedral angles ω of the peptide groups) from a completely random starting conformation within less than 2 hours of CPU time (on an IBM RX-6000 machine)? (2) What is the energy uncertainty above the GM that is admissible (i.e., within which the structure of the ligand will still be very close to the X-ray or NMR structure, within an rms deviation of 1 Å)? (3) How computationally complex is the molecular system thrombin–FPR? By answering these questions, we can determine how well our docking algorithm compares with a standard Monte Carlo approach. To answer the third question, we carried out some extensive docking calculations with the same system using the program DOCKVISION developed by Hart and Read.[14] We then discuss the limits of our approach and the progress that can be expected for finding the native structure of a flexible ligand of the size of a tripeptide with high probability within a reasonable amount of computational time.

Methods

MONTE CARLO MINIMIZATION DOCKING ALGORITHM

Our objective in this study is to test our flexible docking algorithm and, more specifically, to assess the reliability of the program in reaching a well-defined GM. For this purpose, we needed to have a structure of a ligand–protein complex for which the global minimum is known. The easiest way to construct this reference state is to start from the PDB X-ray structure of the complex and minimize its energy. This is based on the assumption that the X-ray structure is close to the GM of the ECEPP/3 force field,[27–31] which is actually the case within 1 kcal/mol accuracy and 1 Å in rmsd. These are the differences that we observed for the ligand before and after energy minimization of the complex, thrombin–FPR, starting from the PDB structure (1 ppb). During this energy minimization, only the dihedral angles of the light and heavy chains of thrombin and the ligand FPR were allowed to vary. No external variables were used for the ligand FPR during the minimization. The ligand FPR was not covalently bound to thrombin in this study. This minimized PDB structure was used as our reference X-ray structure, and all rmsd values are expressed with respect to this structure. If the reference state were the original PDB structure, the only way to test the quality of the results would be to calculate the rmsd between the calculated and the X-ray structure of the ligand. In that case, we could not be sure whether the failure of the docking simulations would arise from errors in the force field or from defects in the docking algorithm, especially in the global optimization method used in the program.

In this study, all the residues were considered in their neutral state so as not to bias the docking simulation. It is well known that, for trypsinlike enzymes, the binding of the ligand into the active site is driven by electrostatic interactions. If the residues had full charges, the side chain of Arg-16 would be forced into the binding pocket, more easily than without charges.

The docking simulations were carried out with a flexible ligand and a rigid receptor. At the beginning of a simulation, a precalculated low-energy conformer of the ligand FPR was located randomly inside a sphere of 5.0-Å radius centered at a point P in the middle of the active site near Ser-195. The coordinates of point P are (3.316, 18.274, 17.381 Å) in the coordinate system of the PDB structure. The docking simulation then consisted of tumbling the ligand inside the active site using a Monte Carlo energy minimization (MCM) approach. For the rigid-body rotation and translation variables, the sampling was carried out uniformly without any biases, whereas the perturbation of all the torsional angles during a single MCM step was calculated with the scaled collective variables algorithm developed by Noguti and Gō.[32] The following energy function was used:

$$E = \lambda_{\text{intra}} E_{\text{intra}} + \lambda_{\text{inter}} E_{\text{inter}} + \lambda_{\text{tors}} E_{\text{tors}} + \lambda_{\text{quat}} E_{\text{quat}} + \lambda_{\text{dist}} E_{\text{dist}} \quad (1)$$

The λs are the weights associated with each energy term. E_{intra} stands for the intramolecular ECEPP/3 (electrostatic and Lennard–Jones) energy of the ligand alone. E_{inter} is the energy between the ligand and the receptor which is calculated from the grid (see below for the values of λ_{intra} and λ_{inter}). E_{tors} is the torsional ECEPP/3 energy. The value of λ_{tors} was always equal to 1.0. E_{quat} is the harmonic penalty function to constrain the norm of the quaternions to unity.[31] Its weight was also always equal to 5000. E_{dist} is a harmonic constraint energy term that is different from zero when the pivot atom of the ligand leaves the sphere of 5-Å radius defined earlier (in the present case, the pivot atom is the C^{α} atom of Pro-2 of

FPR); the associated weight λ_{dist} is 200. The dihedral angles were defined in the forward direction from the C^{α} of the middle residue (Pro-2) and in the backward direction from this C^{α} to the N-terminal group. This forward/backward generation procedure has been shown to increase the efficiency of the energy minimization by a factor of 10 when several molecules are involved.[33]

Before the docking simulations, 20 low-energy conformers of FPR were calculated independently without the presence of the receptor, using short MCM simulations of 300 steps each. The structure of FPR was generated initially as extended. Because the energy of this starting conformation is very high, a simulated annealing protocol was applied to the intramolecular energy. This was realized by varying λ_{intra} continuously from 10^{-5} to 1 at the end of the simulation. The temperature parameter, T, in the Boltzmann factor was maintained constant at 1000 K during all the simulations presented in this study. Changing the λ-factor is equivalent to changing the temperature from $1000/10^{-5}$ to 1000 K in the Boltzmann factor. The step size for the dihedral angles was chosen such that the maximum perturbation was around 30°.

After the generation of the conformers, the docking simulation consisted of two independent stages: (i) the first one corresponded to the tumbling of the ligand inside the active site. This was simulated by an MCM calculation of 1500 steps. For this series of simulations, a maximum of 30° was allowed for the changes in the dihedral angles, 180° for the rigid-body rotation, and 1.0 Å for the translation displacement vector. During the tumbling procedure, the simulated annealing schedule was applied to E_{inter} by varying λ_{inter} from 10^{-5} to 1.0. λ_{inter} was multiplied by 3 after every 10 MCM steps. After 110 MCM steps, the value of λ_{inter} was equal to 1.0, and was kept constant thereafter until the end of the simulation. Because the starting conformation of the ligand already had a low energy, the value of λ_{intra} was maintained equal to 1 throughout the rest of the simulation. The lowest energy structure obtained during the tumbling procedure was then used as the starting point for another MCM simulation. (ii) The second stage consisted of 1000 MCM steps. This last simulation was aimed at energy refinement of the best structure found previously. Therefore, a maximum of 10° perturbations was then allowed for the dihedral angles and for rigid-body rotation and 0.5° for the translation vector. For this second stage of docking simulations, the values of λ_{intra} and λ_{inter} were maintained equal to 1.0.

GRID CALCULATION

The field created by the receptor atoms in the active site region was stored in a 3D grid. This grid was defined as a cube of 56-$Å^3$ volume with a constant spacing of 0.666 Å, with a center located at the point P = (8.0, 13.0, 7.0 Å). The 3D grid was actually split up into 15 different grids: one for the electrostatic field, and 14 grids for the nonbonded energy associated with each of 14 atom types in ECEPP/3. The value of the field at a given point of the active site was interpolated from the 64 surrounding grid points. Interpolation was carried out with the Bezier splines. Calculation of the grids required about 45 minutes on an IBM SP2 RX-6000 (Model 390) computer, and was carried out only once. The 15 grids were stored in binary files of total size 73.7 MB.

Results

MCM DOCKING SIMULATIONS USING BEZIER SPLINES

Twenty independent docking simulations were carried out with a flexible ligand and a rigid receptor. Each independent simulation consisted of three runs: (i) generation of a low-energy conformer of FPR using 300 MCM steps on the ligand alone; (ii) tumbling of the ligand inside the active site of thrombin. The variables were the 17 dihedral angles of FPR plus three translations and four quaternions for rigid-body rotation; (iii) The third MCM simulation of 1000 steps was carried out to refine the lowest energy conformation obtained in the previous simulation. The results for 20 independent simulations are presented in Table I. One series of three MCM simulations required an average of 140, 4900, and 2240 seconds, respectively, which adds up to about 2 hours on an IBM RX-6000 computer.

The results show, first, that all the 20 lowest energy values are close to the GM (−94.6 kcal/mol). They vary between −85 kcal/mol and −95 kcal/mol, with corresponding rmsd values all less than 4.0 Å. Visual inspection of the calculated structures shows that, in all of them, the arginine side chain more or less points toward the pocket of thrombin. Differences in the rmsd result mostly from the position of the Phe side chain and slight differences in the orientation of the overall

TABLE I.

Lowest Energies (in kcal / mol) and Corresponding rmsd Values (Å) of 20 Independent MCM Simulations Using ECEPP / 3-Based Docking Package.[a]

Run No.	MCM Search for Low-Energy Conformers of FPR		MCM Tumbling of Ligand Inside Active Site		MCM Refinement of Ligand Inside Active Site	
	Energy	rmsd	Energy	rmsd	Energy	rmsd
1	−32.9	1.4	−84.7	3.6	**−92.5**	**0.8**
2	−30.5	4.1	−89.2	3.2	−89.2	3.2
3	−31.4	2.0	−86.6	1.3	**−94.6**	**0.6**
4	−33.15	5.0	−85.8	3.7	−88.6	3.2
5	−33.0	4.9	−84.1	3.5	−87.8	3.3
6	−23.3	3.8	−79.0	4.7	−93.4	2.4
7	−32.2	5.3	−83.9	2.5	−86.5	2.5
8	−33.3	5.4	−83.2	3.6	−86.4	3.2
9	−31.4	1.8	−82.7	3.4	−87.0	3.1
10	−30.9	2.3	−82.2	3.6	−87.9	3.3
11	−33.3	4.5	**−94.1**	**0.1**	**−94.6**	**0.5**
12	−31.4	2.0	−86.6	1.5	**−94.6**	**1.0**
13	−31.4	2.2	−88.4	3.1	−89.2	3.1
14	−31.0	4.0	−84.2	4.1	−85.8	4.1
15	−33.0	5.0	−85.1	3.6	−89.2	3.2
16	−32.9	1.4	**−94.6**	**0.5**	**−94.6**	**0.5**
17	−33.1	4.7	−82.7	4.1	−87.2	3.3
18	−31.9	1.1	−83.6	1.5	**−94.6**	**1.0**
19	−35.9	1.7	−83.3	4.1	−85.8	4.1
20	−33.3	5.4	−85.6	3.5	−88.0	3.1

[a] The GM ± 1 kcal / mol is shown in boldface.

structure compared with the X-ray structure. Twenty-five percent (5 of 20) of the simulations reached the GM (−94.6 kcal/mol), which is in fact lower than the energy of the reference state (−94.3 kcal/mol). The five structures at −94.6 kcal/mol are clustered into two different states with rmsd 0.5 and 1.0 Å, respectively. An error of 1-Å rmsd between the calculated and X-ray structures seems to be the level of accuracy one can achieve with an all-atom force field such as ECEPP/3. If such a degree of precision is required, an energy of −94.6 ± 1.0 kcal/mol has to be obtained. A conformation of higher energy by 1.2 kcal/mol compared with the GM already differs from the X-ray by 2.4 Å (see simulation no. 6 in Table I). For energies above the GM ± 1.0 kcal/mol, there is no clear correlation between the energy and the rmsd. These simulations show, therefore, the importance of reaching the GM of the energy function of the ligand–receptor complex. According to our experience, it seems that the energy gap between the GM and the other energy states decreases as molecular flexibility is added to the system—for example, to the receptor (results not shown).

STANDARD MC DOCKING SIMULATION USING THE DOCKVISION PROGRAM

To test the efficiency of our algorithm and how easy or difficult it is to treat the docking of FPR onto thrombin, we compared our ECCEP/3-based docking program with DOCKVISION, developed by Hart and Read.[14] It should be emphasized that the source code was not available to us and, therefore, the DOCKVISION algorithm could not be optimized as was done for our ECEPP/3-based program. Second, to compare the program on exactly the same geometry of the receptor active site, the same reference energy-minimized structure was used as the target molecule in the DOCKVISION program. This reference state might not correspond to the GM of the DOCKVISION potential function. However, it corresponds to a low energy for both the ECEPP/3 and DOCKVISION research force field (see Table II). We may assume that the reference structure is not very far from the GM of the DOCKVISION research force field.

The goal of these tests was first to compare the efficiency of the MCM algorithm with the standard MC algorithm by using the best possible docking

TABLE II.
Lowest Energy Structure (in kcal / mol) and Lowest rmsd Structure (Å) for Ten Independent Docking Simulations Using DOCKVISION.

Simulation No.	Lowest Energy Conformation		Lowest rmsd Conformation	
	Energy	rmsd	Energy	rmsd
X-ray	**−71.9**[a]	**1.1**	**−63.9**	**0.6**
1	−57.0	7.0	−44.0	6.9
2	−54.2	4.7	−49.8	3.4
3	−57.0	9.5	−37.5	6.8
4	−60.4	7.8	−52.5	4.5
5	−58.5	3.8	−49.7	2.7
6	−57.6	9.4	−47.8	4.5
7	−56.1	9.4	−48.2	3.8
8	−58.1	5.4	−56.1	2.2
9	−56.8	7.6	−48.9	3.1
10	−56.9	2.6	−56.9	2.6

[a] The difference of energy scales between the ECEPP/3-based and DOCKVISION programs and between the coarse (Table III) and refined energy mode (Table II) of the DOCKVISION program prevent any comparison between Tables I, II, and III.

package for each of them. For the standard MC algorithm, we expect that the DOCKVISION program will work better than our ECEPP/3 version of the program using the standard MC algorithm. Second, we wanted to be sure that our chosen example, the docking of FPR onto thrombin, was not a trivial one that can be carried out easily with another docking package. DOCKVISION simulations were carried out on an Indigo-2 Silicon Graphics workstation, which is 1.4 times faster than the IBM RX-6000 computer.

Two types of test were carried out with DOCKVISION. The first series of tests consisted of docking the FPR ligand into the rigid receptor starting from the first ten low-energy conformers found after the first stage of our procedure (columns 2 and 3 in Table I). The starting conformation, orientation, and position of the ligand FPR inside the active site were thus exactly the same as the starting configuration of our tumbling procedure (that had led to columns 4 and 5 of Table I). But, in this test, our tumbling and energy refinement procedures were replaced by a sequence of five independent standard MC runs with decreasing temperature, T, in which the last accepted conformation of each run was used as the starting point of the following MC run. The five runs consisted of 500, 500, 500, 500, and 1000 MC steps each. The corresponding temperature parameters were 300, 200, 100, 50, and 10 K. The maximum perturbations for the dihedral angles and rigid-body rotations were 30° for the first three runs, 10° for the fourth, and 6° for the last run. The maximum perturbation for the translation vector was 1 Å for the first four runs and 0.3 Å for the last run. At each MC step, a quick evaluation of the energy of interaction between the ligand and the receptor was carried out with a precalculated 3D grid. If the energy was above a certain threshold, predefined by the user (−20 kcal/mol in this case), the energy was re-evaluated using the complete N^2 interactions between the ligand and the receptor. To speed up the calculations, a cutoff distance of 8 Å was used. The internal conformational energy term of the ligand was not turned off, and was thus taken into account. For each of the ten independent docking simulations, the five MC runs were repeated 1500 times using different series of random numbers. Every structure with a negative energy found during the last MC run of the simulated annealing procedure (1000 MC steps at 10 K) for each of the 1500 independent simulations was stored on a graph showing the rmsd vs. energy. The lowest energy point and its corresponding rmsd among all those recorded from the graph are listed in Table II (columns 2 and 3). The lowest rmsd and its corresponding energy were also recorded (last two columns of Table II). The simulated annealing schedule was designed in such a way that the computational time was more or less equivalent to the one used for ECEPP/3-based simulation. One single simulation required about 1.5 hours on an Indigo-2 Silicon Graphics workstation, which is about 1.4 faster than the IBM RX-6000 computer.

The result shows that none of the lowest energy conformations of FPR was close to the X-ray reference energy (−71.9 kcal/mol). The calculated lowest energies fluctuate between −54.2 and −60.4 kcal/mol. Usually, a refinement procedure could be applied to these low-energy conformers to sort them out according to a more sophisticated free energy function. Columns 4 and 5 of Table II present the lowest rmsd conformations found during the sampling process. Unless local energy minimization is used on the resulting set of structures, the lowest rmsd values show the best ranking that can be obtained with a given energy evaluation procedure. In this test, the lowest rmsd value is 2.2 Å (simulation no. 8 of Table II). The corresponding energy is about 8 kcal/mol above the GM.

A second type of test was carried out to sample the conformational space of the system more ex-

tensively. Three independent simulations were carried out with the same simulated annealing schedule (five successive runs with decreasing temperature). The difference from the previous tests is that, in this type of test, the structures of 1000 conformers of FPR were previously precalculated in the absence of the receptor, and the five independent runs were carried out 3000 times with different series of random numbers, starting randomly from one of the 1000 conformers. This corresponds to switching off the ''Refinement'' option of the DOCKVISION program. The ligand conformer was located randomly inside the active site within a sphere of 5-Å radius instead of starting from our predefined conformation, orientation, and location (i.e., the starting configurations of the first ten rows of columns 4 and 5 of Table I), as in the first test. The center of the sphere is the same as the one defined in our ECEPP/3-based simulations. The results of this second test are presented in Table III. One set of 3000 simulated annealing runs took 3 hours of CPU time on the Indigo-2. The results show that none of the calculated lowest energy conformations of FPR were close to the X-ray structure. The lowest rmsd was 3.7 Å (simulation no. 1 of Table III). These results suggest that the docking of FPR onto thrombin cannot be carried out easily with standard MC algorithms. We tried to run a standard MC simulation using the ECEPP/3-based package, with and without the Noguti–Gō protocol, and without energy minimization; however, we had no success. In this respect, the standard MC approach of DOCKVISION was far more efficient than our simple multiple-start standard MC tests carried out with the ECEPP/3-based docking program.

TABLE III.
Lowest Energies (kcal / mol) with Corresponding rmsd Values (Å) and Lowest rmsd Values with Corresponding Energy Values for Second Test that Consisted of Three Independent Docking Simulations Using DOCKVISION.

Simulation No.	Lowest Energy eConformation		Lowest rmsd Conformation	
	Energy	rmsd	Energy	rmsd
1	−12.0	3.7	4.8	1.7
2	−16.0	5.5	0.5	1.6
3	−8.0	7.0	2.0	1.9

Discussion and Conclusions

The comparison tests made with the FPR–thrombin complex show that the probability of reaching the GM of an all-atom potential function such as ECEPP/3 is much higher when local energy minimization is used during the MC sampling. By comparing the energy and rmsd values listed in Table I, attainment of the GM within 1.0 kcal/mol is required to ensure that the corresponding structure is less than 1.0-Å rmsd from the experimental structure. When the receptor is rigid and the ligand flexible, the lowest value of the ECEPP/3 energy function corresponds to the native structure of the ligand. This might no longer be the case when flexibility is added to the receptor (results not shown). For such a system, the interaction between the receptor and the solvent will undoubtedly play a major role. Consequently, a more exact physical model for the ligand–receptor interaction (including solvation) is necessary for the native structure to correspond to the GM of this more sophisticated energy function. When the receptor is rigid, the ECEPP/3 function without solvation was good enough to find the X-ray ligand conformation and to distinguish between the native and nonnative structure of the ligand.

The introduction of a fast local energy minimization using Bezier splines is the major improvement in the efficiency of the docking simulations. On the other hand, the advantage of the scaled collective variables MC over the standard MC depends on the type of complex being simulated. Usually, the relative efficiency increases with the size of the flexible part introduced in the system. In this study, the algorithm consists of tumbling the ligand inside the active site of the receptor. Therefore, the external variables of the ligand were sampled randomly with a uniform probability distribution without taking into account the information in the hessian. When the ligand is larger (five residues or more), it is more appropriate to consider the hessian eigenvectors for the generation of all the variables, internal and external, as well. This option, of course, is present in the program.

Our comparison with the DOCKVISION program was not intended to test the quality of the DOCKVISION program, but rather to assess the difficulty of our single chosen docking example. We wanted to know if the docking of FPR onto thrombin was trivial. If trivial, there is no relevance to testing a

new docking algorithm on an easy system for which good results can be obtained, regardless of the performance of the docking algorithm. Also, in the comparison between the SCV MC and the standard MC algorithm, it is more fair to use a program for which the standard MC algorithm has been well tuned and optimized. Our own tests of standard MC simulations were unsuccessful because we did not spend adequate time tuning the algorithm.

The conclusion from the DOCKVISION simulation is that this docking example is in fact nontrivial and that the introduction of a local energy minimization after each MC step leads to a major improvement in efficiency in reaching the GM for docking simulations. This solves the problem of the presence of false positive results in the list of candidates of low-energy conformers. The generation of the list therefore becomes relevant only if we expect several binding modes for the given ligand.

It is important to note that better annealing schedules than those that we used for the tests with DOCKVISION can be found by more experienced users. One way to speed up the simulation time of DOCKVISION is to switch off the energy term E_{intra} of the ligand. Also, better results can always be obtained by carrying out additional energy refinement stages.

A flexible docking program using an MCM algorithm has recently been developed.[34] Molecular flexibility was assigned to both the ligand and the receptor. A distance cutoff of 8 Å was used to increase the speed of the local energy minimization. Because of possible atomic clashes between the random starting conformation of the ligand and receptor, infinite energy values might arise during the initial steps of the energy minimization. To avoid this problem, the original Lennard–Jones energy function was modified[34] in such a way that, at zero distance between two atoms, the energy was not infinite. This problem does not arise when the interaction energy between the ligand and receptor is interpolated from the grid points using Bezier splines, as was done in our procedure. Therefore, the starting conformation of the ligand with atomic clashes with the receptor can easily be energy minimized. Besides the advantage of the Bezier splines in speeding up the calculation of interaction energy, they also have the ability to smooth out the potential energy surface.[21] The energy barriers are thus lower and the probability of reaching the GM is thus increased as compared with the same MCM algorithm without the grid and the Bezier splines technique.

With the MCM algorithm and Bezier splines grid, the probability of obtaining the GM within 2 hours of computation time was 25%. This is significant but is still insufficient for our goal of docking a large number of inhibitors in a relatively short amount of time. Progress to improve this percentage can be directed toward the implementation of a more powerful global optimization method. Some of these methods, such as the diffusion equation method,[35] the distance scaling method,[36] and the conformational space annealing method,[22] have been shown to be efficient for finding the GM of polypeptides of 20 amino acids or more. Another factor that can improve the success of the semiflexible docking simulation would be the increased speed of new-generation computers. We may expect a factor of ten speed-up from a more powerful computer. For the time being, there is actually an easy ''trick'' to make the flexible docking algorithm nearly as fast as the rigid docking algorithm. This trick, which is used in DOCKVISION, is to switch off the energy term E_{intra} of the flexible ligand. The only interatomic energy term is E_{inter}, between the ligand and the receptor; it is evaluated from the grid points, which is very fast. We, therefore, may not be too far from calculating the structure of a small flexible ligand bound to its receptor in such a way that a single simulation can reach the GM with high probability. This can be a good starting point for screening a large number of drug molecules.

Acknowledgment

We thank T. N. Hart for kindly providing the DOCKVISION program and for the helpful guidance in its use. We thank the Association Française pour la Recherche Thérapeutique and the Centre National de la Recherche Scientifique IMABIO for fellowship support to J.-Y. T. This work was supported by NSF grant No. MCB95-13167, by NIH research grants Nos. GM-14312 and HL-30616, and by the Cornell Biotechnology Center. The calculations were carried out on the IBM SP2 supercomputer at the Cornell Theory Center, which receives funding from Cornell University, the State of New York, the National Center for Research Resources at the National Institutes of Health (NIH), and members of the Center's Corporate Partnership Program.

References

1. DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. J Med Chem 1988, 31, 722–729.
2. Shoichet, B. K.; Kuntz, I. D. J Mol Biol 1991, 221, 327–346.
3. Bacon, D. J.; Moult, J. J Mol Biol 1992, 225, 849–858.
4. Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. Proc Natl Acad Sci USA 1992, 89, 2195–2199.
5. Kasinos, N.; Lilley, G. A.; Subbarao, N.; Haneef, I. Prot Eng 1992, 5, 69–75.
6. Shoichet, B. K.; Bodian, D. L.; Kuntz, I. D. J Comput Chem 1992, 13, 380–397.
7. Shoichet, B. K.; Kuntz, I. D. Prot Eng 1993, 6, 723–732.
8. Helmer-Citterich, M.; Tramontano, A. J Mol Biol 1994, 235, 1021–1031.
9. Vakser, I. A.; Aflalo, C. Prot Struct Function Genet 1994, 20, 320–329.
10. Meyer, M.; Wilson, P.; Schomburg, D. J Mol Biol 1996, 264, 199–210.
11. Sobolev, V.; Wade, R. C.; Vriend, G.; Edelman, M. Prot Struct Function Genet 1996, 25, 120–129.
12. Friedman, J. M. Prot Eng 1997, 10, 851–863.
13. Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. E. J Mol Biol 1997, 272, 106–120.
14. Hart, T. N.; Read, R. J. Prot Struct Function Genet 1992, 13, 206–222.
15. Ewing, T. J. A.; Kuntz, I. D. J Comput Chem 1997, 18, 1175–1189.
16. Caflisch, A.; Niederer, P.; Anliker, M. Prot Struct Function Genet 1992, 13, 223–230.
17. Guida, W. C.; Bohacek, R. S.; Erion, M. D. J Comput Chem 1992, 13, 214–228.
18. Abagyan, R.; Totrov, M.; Kuznetsov, D. J Comput Chem 1994, 15, 488–506.
19. Li, Z.; Scheraga, H. A. Proc Natl Acad Sci USA 1987, 84, 6611–6615.
20. Li, Z.; Scheraga, H. A. J Mol Struct (Theochem) 1988, 179, 333–352.
21. Trosset, J-Y.; Scheraga, H. A. Proc Natl Acad Sci USA 1998, 95, 8011–8015.
22. Lee, J.; Scheraga, H. A.; Rackovsky, S. J Comput Chem 1997, 18, 1222–1232.
23. Oberlin, Jr., D.; Scheraga, H. A. J Comput Chem 1998, 19, 71–85.
24. Luty, B. A.; Wasserman, Z. R.; Stouten, P. F. W.; Hodge, C. N.; Zacharias, M.; McCammon, J. A. J Comput Chem 1995, 16, 454–464.
25. Bode, W.; Mayr, I.; Baumann, U.; Huber, R.; Stone, S. R.; Hofsteenge, J. EMBO J 1989, 8, 3467–3475.
26. Bode, W.; Turk, D.; Karshikov, A. Prot Sci 1992, 1, 426–471.
27. Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. J Phys Chem 1975, 79, 2361–2381.
28. Némethy, G.; Pottle, M. S.; Scheraga, H. A. J Phys Chem 1983, 87, 1883–1887.
29. Sippl, M. J.; Némethy, G.; Scheraga, H. A. J Phys Chem 1984, 88, 6231–6233.
30. Némethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. J Phys Chem 1992, 96, 6472–6484.
31. Trosset, J-Y.; Scheraga, H. A. J Comput Chem (in press).
32. Noguti, T.; Gō, N. Biopolymers 1985, 24, 527–546.
33. Gibson, K. D., Scheraga, H. A. J Comput Chem 1994, 15, 1403–1413.
34. Apostolakis, J.; Plückthun, A.; Caflisch, A. J Comput Chem 1998, 19, 21–37.
35. Kostrowicki, J.; Scheraga, H. A. DIMACS Series Discrete Math Theor Comput Sci 1996, 23, 123–132.
36. Pillardy, J.; Piela, L. J Comput Chem 1997, 18, 2040–2049.